Isohelicity and Phasing in Drug−DNA Sequence Recognition: Crystal Structure of a Tris(benzimidazole)−Oligonucleotide Complex†

George R. Clark,‡ Emily J. Gray, and Stephen Neidle*

The CRC Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, U.K.

Yu-Hua Li and Werner Leupin§

Preclinical Research Pharma Gene Technologies, F. Hoffman-La Roche Ltd., CH-4002 Basel, Switzerland

Received February 21, 1996; Revised Manuscript Received May 20, 1996⊗

ABSTRACT: The crystal structure is reported of a tris(benzimidazole) analogue of the minor-groove drug Hoechst 33258 bound to the sequence d(CGCAAATTTGCG)$_2$. The structure has been refined to an *R* factor of 17.4% at a resolution of 2.2 Å. The ligand covers ~7$^1/_2$ base pairs, including the 5′-AAATTT central sequence. This has an exceptionally narrow minor-groove width, together with high propeller twists for individual base pairs. The ligand has a highly twisted structure, with an overall twist of 50° between aromatic rings. All three benzimidazole subunits are in register with the DNA, and there is a symmetric group of six hydrogen bonds between ligand and A·T base-pair edges. By contrast, the ligand does not show an optimal isohelical fit to the DNA. The correct phasing of drug and DNA base pairs is ensured by a number of changes to the DNA such that the central 5′-AAATTT region is slightly unwound relative to the structures of other noncovalent minor-groove drug complexes.

A variety of low molecular weight molecules, typified by netropsin, distamycin, berenil, pentamidine, and Hoechst 33258, interact noncovalently with AT-rich sequences in the minor groove of double-stranded DNA (Dervan, 1986; Zimmer & Wähnert, 1986; Kopka & Larsen, 1992; Abu-Daya et al., 1995). They and many of their derivatives can possess a wide range of potential therapeutic properties, ranging from antitumor to antiviral, antimicrobial, antimalarial, and antifungal activity [see, for example, Tidwell et al. (1990), Broggini et al. (1991), Prakash et al. (1991), Gravatt et al. (1994), and Boykin et al. (1995)]. The relationships between DNA-binding ability in vitro and these biological activities in vivo are necessarily complex, although there is increasing evidence that these drug molecules may compete with transcription factor or homeodomain binding to DNA sequences (Dorn et al., 1992; Chiang et al., 1994; Welch et al., 1994; Henderson & Hurley, 1995; Bellorini et al., 1995). Several minor-groove binders have also been shown to interfere with the functioning of DNA topoisomerases I or II in cells (Beerman et al., 1992; Chen et al., 1993).

A number of X-ray crystallographic studies have been reported for these minor-groove drugs complexed with dodecanucleotides [for example, Kopka et al. (1985), Coll et al. (1987), Brown et al. (1990, 1992), Edwards et al. (1992b), Tabernero et al. (1993), and Nunn and Neidle, (1995)]. All show drug bound in the narrow minor groove of AT-rich sequences, with extensive van der Waals contacts with the floor and walls of the groove. Hydrogen bonding from drug to sites on base-pair edges has been frequently, though not invariably, observed. These crystal structures have been used as the starting points for the rational design of molecules capable of recognizing longer AT sequences (Youngquist & Dervan, 1985, 1987; Dervan, 1986) as well as DNA sites containing other than purely AT base pairs [see, for recent examples, Mrksich et al. (1994), Singh et al. (1994), Chen and Lown (1994)]. A key factor in these studies has been the ability of designed molecules to fit the minor groove in an effective manner, such that the convex surface of the groove floor is complemented in shape by the concave inner-facing surface of the ligand. This property ("isohelicity") has been quantified in terms of ligand radius of curvature and ranking orders determined for series of building blocks (Goodsell & Dickerson, 1986; Zasedatelev, 1991; Cory et al., 1992) on the basis of complementarity to the minor groove of canonical B-DNA. Studies with molecules composed of multiple pyrrolecarboxamide subunits (Youngquist & Dervan, 1985) showed that linking more than about seven such units resulted in decreased DNA affinity, presumably on account of the ligand's hydrogen-bonding groups no longer being in register ("in phase") with those of the DNA.

The bis(benzimidazole) compound Hoechst 33258 (**1**, Figure 1) has received particular attention as a starting point for the design of new potential anticancer drugs. The parent compound has cytotoxic activity (Chen et al., 1993) and has been in phase I/II clinical trials against pancreatic carcinomas (Patel et al., 1991). A number of Hoechst analogues have been synthesized and evaluated for biological activity [see, for example, Bathini et al. (1990), Beerman et al. (1992), Wong et al. (1994), and Sun et al. (1994)]. In many, though not all, instances, DNA affinity correlates with biological response. Hoechst 33258 has been found by footprinting

† Studies at the Institute of Cancer Research were supported by a program grant to S.N. from the Cancer Research Campaign.

* Address correspondence to this author. Tel/Fax: (44)181 643 1675. Email: steve@iris5.icr.ac.uk.

‡ Permanent address: Chemistry Department, University of Auckland, Auckland, New Zealand.

§ Present address: Department of Physical Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0006-2960(96)00421-7 CCC: $12.00

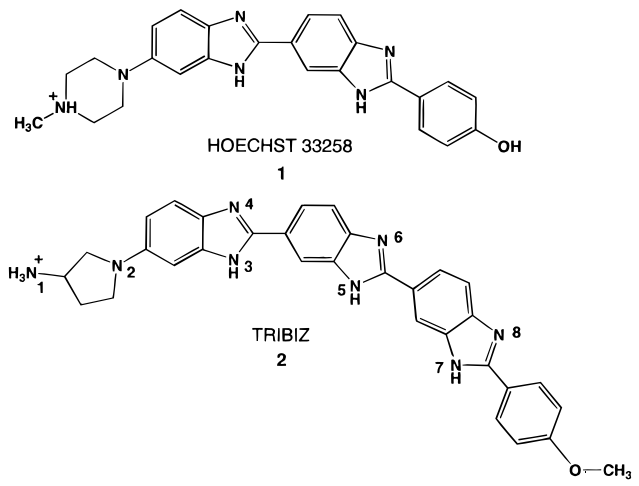

FIGURE 1: Structural formulas of Hoechst 33258 (**1**) and TRIBIZ (**2**).

methods (Harshman & Dervan, 1985) to bind to DNA sites 4−5 base pairs in length, with the ability to accommodate a GC base pair at the 5′ end of the site. Several crystallographic analyses have shown the drug bound within runs of AT base pairs in dodecamers, with the involvement of hydrogen bonds from the inner-facing nitrogen atom of each benzimidazole ring to AT base pairs (Pjura et al., 1987; Teng et al., 1988; Quintana et al., 1991; Carrondo et al., 1989; Spink et al., 1994; Vega et al., 1994). More recently, we have shown (Wood et al., 1995) that replacement of the bulky piperazine ring in Hoechst 33258 by the smaller imidazole group results in changes in groove width that reflect this decrease in steric bulk.

In addition to the X-ray crystallographic studies of complexes between oligonucleotides and Hoechst 33258 (see above), several NMR studies have been published [see Fede et al. (1991, 1993) and references therein]. In one case (Fede et al., 1993) the three-dimensional structure of the 1:1 complex between Hoechst 33258 and d(GTGGAATTC-CAC)$_2$ has been determined. The results of this study have been used as a basis for molecular modeling studies and subsequent synthesis of tris(benzimidazole) derivatives (Ji et al., in preparation) such as **2** (called "TRIBIZ"; Figure 1). TRIBIZ exhibits DNA-binding activity and shows some antibacterial and good antimalarial activity (Ji et al., in preparation).

We report here on the crystal structure of the 1:1 complex between the dodecanucleotide duplex d(CGCAAATTTGCG)$_2$ and TRIBIZ. This structure enables us to examine (i) the ability of this molecule to be isohelical with DNA along its length, (ii) the extent that ligand hydrogen bonding to DNA remains in phase with the helical twist of the base pairs along the binding site, and (iii) the conformational adaptability of DNA when confronted with such a ligand. TRIBIZ would be expected to span a binding site of $\sim 7^1/_2$ base pairs ($\sim^3/_4$ of a complete turn of B-DNA) on the basis of its likely ~23 Å end to end length and so is comparable in size with binding sites found for several minor-groove-binding proteins (Feng et al., 1994; Werner et al., 1995). DNA footprinting studies have indicated that it has a protected site size of at least 6 base pairs (Ji et al., in preparation).

MATERIALS AND METHODS

Synthesis and Crystallization. The DNA dodecamer d(CGCAAATTTGCG)$_2$ was purchased from the Oswel DNA Service (University of Edinburgh) and annealed before use. The TRIBIZ ligand was prepared as the hydrochloride salt, as reported elsewhere (Ji et al., in preparation).

The ligand−DNA complex was grown from hanging drops at 286 K as light yellow elongated prismatic-shaped crystals. The crystal used for data collection was grown from a drop containing 3 μL of 35% 2-methylpentane-2,4-diol, 2 μL of 5 mM TRIBIZ, 3 μL of 200 mM $MgCl_2$, 1 μL of 2.5 mM spermine, and 2 μL of 5 mM dodecamer equilibrated against a reservoir containing 1 mL of 40% 2-methyl-2,4-pentanediol. The DNA solution was prepared using 30 mM sodium cacodylate buffer at pH 7.0. The crystal employed for the X-ray study was obtained after about 4 months.

Crystallographic Data Collection. The crystal used for data collection was of approximate dimensions 0.44 × 0.12 × 0.04 mm and was mounted in a 0.5-mm Lindemann glass capillary with a small amount of mother liquor. Intensity data were collected at 287 K using a Siemens-Xentronics multiwire area detector with a rotating anode X-ray generator (40 mA, 75 kV) and a graphite monochromator. A crystal-to-detector distance of 10 cm and swing angle of 15° were used to collect data to a maximum possible resolution of 2.2 Å. Data were collected with χ set at 45°, while the crystal was rotated through 100° in ω at ϕ values of 0° and 60°; 180 s frames were recorded every 0.20° step. The crystal did not suffer any observable decay during the data collection. Data processing was carried out using the program package XENGEN, version 1.3. After merging, the data comprised 3317 of the possible 3594 unique reflections to 2.2 Å (92.3%), with the outermost shell (2.35−2.20 Å) having 57% observed reflections. The overall merging R value is 8.3% for the data to 2.2 Å.

Structure Determination and Refinement. The unit cell dimensions of the crystal are a = 24.70 Å, b = 40.82 Å, and c = 64.90 Å in the orthorhombic space group $P2_12_12_1$. This cell is close to those reported by us for the native d(CGCAAATTTGCG)$_2$ dodecamer (Edwards et al., 1992a; a = 24.87 Å, b = 40.90 Å, and c = 65.64 Å) and other groove-bound dodecamer−drug complexes, suggesting that the present crystals are isomorphous with them. The dodecamer coordinates used as a starting model for the structure refinement were those of the propamidine−d(CG-CAAATTTGCG)$_2$ complex (Nunn & Neidle, 1995; a = 24.78 Å, b = 41.16 Å, and c = 65.51 Å).

The crystallographic refinement was carried out using the program X-PLOR, version 3.1 (Brünger et al., 1987). Rigid body refinement of the DNA molecule as one constrained group was performed with the resolution range of the data increased from 8.0−4.0 Å (543 reflections) to 8.0−3.5 Å (833 reflections). The R value was 31.5% at this point. The molecule was then divided into 24 rigid groups comprising individual nucleotide groups. The resolution range was gradually increased from 8.0−3.5 Å to 8.0−2.2 Å (2534 reflections), the R factor then being 29.2%. Electron density maps were calculated and displayed using the graphics package TOM/FRODO, version 3.2.

The DNA molecule fitted the density well, and a long continuous lobe of density was clearly visible in the minor groove. This lobe displayed sufficient structure to enable the correct orientation and positioning of the drug molecule to be unambiguously deduced at the outset. We based our conclusions on the following arguements: (i) the length of the continuous lobe exactly matched the expected length of the drug molecule; hence, we deduced that there was no

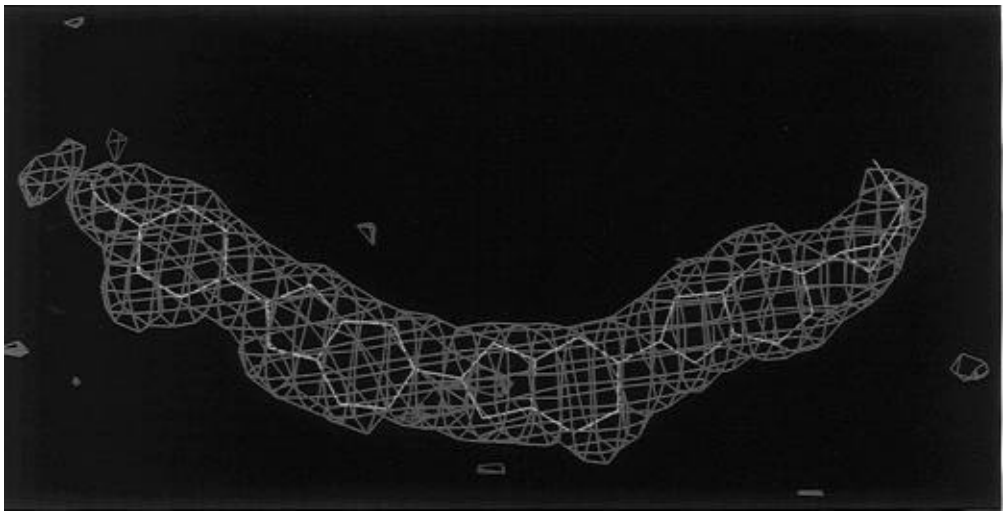

FIGURE 2: Omit map showing the electron density in the minor groove with the final structure of the TRIBIZ molecule superimposed on it. The map has been contoured at the 2.5σ level and shows all electron density within 3 Å of the molecule.

mobility of the drug along the groove. (ii) Assuming no mobility, the drug could lie in one of two orientations or be a statistical mixture of both. The two directions are crystallographically distinct (though chemically equivalent). The lobe of electron density displayed distinct regions of narrowing between the component segments of the drug, indicating that the drug position was not a mixture of both orientations. (iii) The three segments of the density were each assigned to a benzimidazole group; these were each wider at one end than the other, in a regular pattern, giving a clear indication in each case of which end corresponded to a six-membered ring and which corresponded to a five-membered ring of a benzimidazole group. This also supported the deduction that the density did not represent a mixture of both possible orientations. (iv) The shape of the density lobe differed at the two ends, one being noticeably less broad in one dimension than the other. This indicated that the end with this thin cross section contained the methoxyphenyl group, and the more bulbous end contained the substituted pyrrolidine group. (v) These last two conclusions were internally consistent with the known structure of the drug molecule, with the methoxyphenyl group attached to a six- rather than a five-membered ring of a benzimidazole group. Many of these features are shown in Figure 2, where we have depicted the best overall view of the drug and electron density lobe in a single map showing that they are in a continuous spiral covering nearly three-quarters of a turn of the DNA.

A structure for a TRIBIZ molecule was constructed using computer graphics (Hyperchem, version 4) by modifying a Hoechst 33258 molecule by inclusion of an additional benzimidazole moiety and changes of the terminal groups. The stepwise twist of the benzimidazoles observed in other complexes (about 19° per benzimidazole) was retained. The trial TRIBIZ model closely fitted the electron density. Electrostatic charges for the TRIBIZ molecule were calculated with Hyperchem using modified neglect of differential overlap (MNDO) wave functions, and the force-field parameters were interpolated from previous studies in this laboratory (Wood et al., 1995). Planar restraints were applied to each individual aromatic ring system.

The drug molecule was included, and the refinement repeated stepwise from low-resolution to high-resolution data. The *R* factor at this point was 22.4%. As a final check on the drug orientation, we tested an alternative model in which the drug was inserted in the opposite orientation, even though it did not fit the observed difference electron density lobe so well. Refinement following the same procedure as above resulted in an *R* factor of 22.5%, essentially indistinguishable from before. We therefore concluded that the assignment of orientation must rest on the arguments detailed above, together with considerations of solvent structure (see below). We therefore returned to the original orientation for further refinement.

Solvent positions were included and assigned as water molecules. The criteria for acceptance of solvent were proximity to the DNA−drug complex (within 6 Å), peak height $>3\sigma$ in difference maps, potential hydrogen-bonding neighbors (2.2−3.4 Å), and sensible thermal parameters.

At the end of the refinement a total of 71 water molecules have been included, and the final *R* value is 17.4% for all observed data in the range 8.0−2.2 Å. The root-mean-square deviations from target values are 0.013 Å for bond lengths and 2.03° for bond angles. The X-PLOR refinement and *R* factor calculations used data with $F > 2\sigma(F)$ (2534 reflections). The mean thermal parameters are 27.8 Å^2 for phosphates, 21.0 Å^2 for sugars, 13.8 Å^2 for bases, 16.2 Å^2 for atoms of the drug, and 42.4 Å^2 for water molecules. Atomic coordinates have been deposited in the Nucleic Acid and Brookhaven Protein Data Banks (file name GDL039).

There are three other features of the refined structure which have a bearing on the selection of the orientation of the drug in the minor groove. (i) The hydrogen-bonding network. As described in more detail later, the exocyclic pyrrolidine substituent $-NH_3^+$ makes a chemically compact set of hydrogen bonds to three water molecules which occupy the void at the 3′ end between the pyrrolidine group at the the end of the drug and the sugar ring from a neighboring dodecamer molecule which projects into the end of the minor groove. (This is the crystal packing effect, which is a common feature of orthorhombic dodecamer crystal structures and which is a factor in the crystallographic inequivalence of the "up" and "down" directions of the DNA duplex.) The three water molecules fit neatly into the void and hydrogen bond to nucleotides on both sides of the minor groove and to the inserted sugar moiety (Figure 7). There is no hydrogen bonding at the methoxyphenyl end of the drug, solely van der Waals contacts between the methoxy

carbon atom and sugar oxygen atoms. The alternative drug orientation would place the hydrophobic methoxyphenyl ring in unacceptably close contact with the three water molecules described above. (ii) The refined temperature factors are relatively low for the sugars, bases, and drug molecule, and are somewhat higher for the alternative orientation, with the drug atoms having a mean temperature factor of 19.6 Å^2. (iii) The compact nature of the binding of the methoxyphenyl ring into the narrow region of the minor groove shows that there would not be room for a bulky pyrrolidine group to penetrate as deeply into the groove without major distortion of the DNA in this region.

RESULTS

Overall Features of the Structure. The overall structure of the d(CGCAAATTTGCG)$_2$ duplex is similar to that found in other isomorphous dodecamers and dodecamer−drug complexes. The helix is right-handed B-DNA. There are 10.1 base pairs per turn of the helix, with a mean helical twist angle between pairs of 35.9°. The mean rise per base pair is 3.36 Å. The crystallographic asymmetric unit consists of two chemically equivalent self-complementary dodecanucleotide strands in an antiparallel duplex. The nucleotides are numbered C1 through G12 starting at the 5′ end for one strand and C13 through G24 for the other. The TRIBIZ drug molecule is labeled residue 25, and the 71 water molecules are labeled W26 to W96.

In the crystal the duplexes pack along the crystallographic *c* axis with the sugar−phosphate backbones of neighboring molecules inserting into each end of the minor groove, in the manner documented for other dodecamer structures. These intermolecular interactions involve the terminal two nucleotides at each end of a duplex, thus leaving the central 8 base pairs free of any direct crystal packing forces. The two strands of the duplex are crystallographically distinct.

The structure of the complex between TRIBIZ and the DNA dodecamer d(CGCAAATTTGCG)$_2$ is shown in Figure 3. The drug is bound in the extended AT region of the minor groove of the duplex and occupies a unique position with no evidence of disorder along the sequence. The binding site extends almost symmetrically beyond the 5′-AAATTT sequence, on both 5′ and 3′ sides, and the far ends of the TRIBIZ molecule are close to the GC base pairs such that the total ligand binding site size is ~7$^1/_2$ base pairs. The TRIBIZ molecule thus spans a longer sequence of DNA than any other groove-binding drug so far studied by X-ray crystallography. The length of the TRIBIZ molecule is such that it must undergo a twist at each junction of its five molecular subunits in order to follow the helical curvature of the minor groove. These twists are in addition to the two-dimensional crescent shape indicated in Figure 1. The dihedral angles between the aromatic rings of adjacent benzimidazole rings are 23° (rings 1/2) and 18° (rings 2/3), and the terminal benzene ring is twisted by a further 9° with respect to benzimidazole ring 3. Thus the overall twist of the aromatic rings in TRIBIZ is ~50° (Figure 3). The angle between the best plane through the pyrrolidine ring atoms and benzimidazole ring 1 is 74°.

Drug−DNA Contacts. All four aromatic ring subunits of TRIBIZ are in van der Waals contact with the floor of the minor groove, with one carbon atom of each of the four benzene rings being close to C2 of an adenine base. These distances are as follows: C27 (methoxyphenyl ring 4)−C2A17, 3.5 Å; C20 (subunit 3)−C2A18, 3.6 Å; C13 (subunit 2)−C2A6, 4.0 Å; C6 (subunit 1)−C2A5, 4.1 Å. There is a clear pattern of increase in these nonbonded distances toward the 3′ end of the binding site, which shows that the TRIBIZ molecule does not penetrate the full depth of the minor groove at the 3′ end.

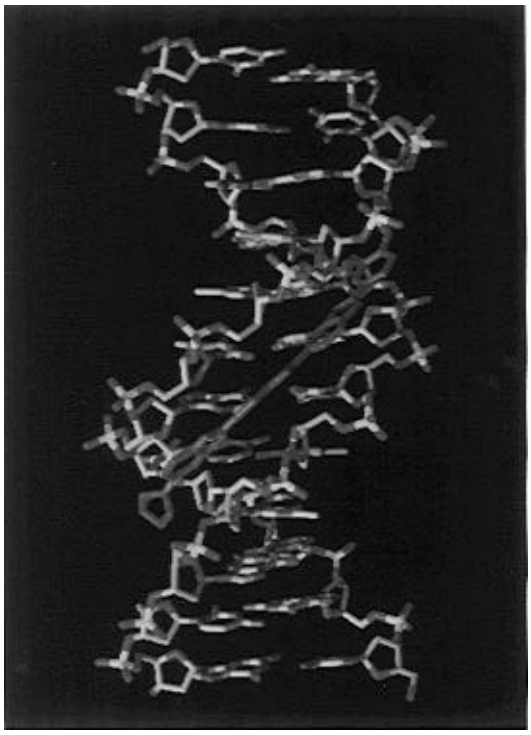

FIGURE 3: Stick representation of the TRIBIZ−d(CGCAAATTTGCG)$_2$ complex, with the carbon atoms of the TRIBIZ molecule colored green. The twists between the benzimidazole subunits are clearly visible. Produced with the MIDAS PLUS program (Ferrin et al., 1988; Huang et al., 1991).

The binding of the TRIBIZ molecule into the minor groove involves a number of hydrogen-bonding interactions. There is a symmetrical sequence of three bifurcated pairs of hydrogen bonds between the inner-facing N3, N5, and N7 atoms of the three benzimidazole groups, and O2 of thymines and N3 of adenine (for two out of the three sets) in the floor of the groove (Figures 4 and 5 and Table 1), for three consecutive AT base pairs. Each of these pairs of hydrogen bonds is not to the bases of an individual base pair but to a staggered pair of bases. The geometry of these hydrogen-bonding arrangements suggests that the donor imidazole NH hydrogen atoms are more or less equally shared between each acceptor in a staggered pair of bases. By contrast, the two pyrrolidine nitrogen atoms are not involved in any hydrogen bonds to the DNA, even though at least one of them is likely to be protonated. The exocyclic N1 atom hydrogen bonds to a water molecule (see later).

There are numerous close van der Waals contacts between the aromatic rings of the drug and the walls of the minor groove, involving mainly the hydrogens attached to atoms C1′, C4′, and C5′ on the phosphodiester backbone; these tend to be directly oriented toward the bound drug. These intermolecular interactions are especially significant in the AT-rich region of the duplex, where the groove is narrow. Groove width (defined in terms of interstrand C4′ distances) is shown in Figure 6, in comparison with that found for several other drug−d(CGCAAATTTGCG)$_2$ complexes. The present structure is characterized by an exceptionally narrow

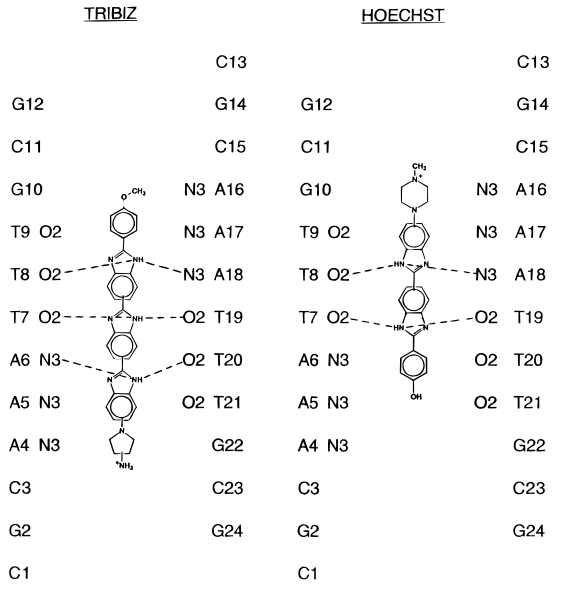

FIGURE 4: Schematic representations of the drug−DNA hydrogen bonding (i) in the TRIBIZ complex and (ii) in the Hoechst complex with the same DNA sequence (Spink et al., 1994).

groove width (3.4−3.6 Å) that extends over most of the AT tract, more so than for the native structure (Edwards et al., 1992a) or for any of the noncovalent minor-groove drug−DNA complexes examined to date. The TRIBIZ pyrrolidine group lies in the wider GC minor groove region at the 3′ end of the binding site, on account of the bulky nature of this group. Atoms of the pyrrolidine group are involved in a number of close van der Waals contacts with atoms on both DNA strands. For strand 1, contacts are made with C4′ and C5′ atoms from Ade6 and the O4′ ribose sugar atom of Gua22 and, for strand 2, with C1′ of Thy21 together with C4′ and C5′ atoms from Gua22. The closest of these contacts is 25C3−22C5′ (3.0 Å). The amino substituent on the pyrrolidine group points into the groove, contributing to these van der Waals contacts. The phenyl ring at the opposite end of the TRIBIZ molecule lies at the end of the narrow minor groove region; its methoxy substituent points into the groove, thus increasing the groove width in this region. The methyl group is in close contact with backbone atoms of strand 1 (25C32−10O4′ 3.4 Å; 25C32−11O4′ 3.6 Å). The outermost contacts with the DNA duplex are the close van der Waals approaches described above between atoms of the aliphatic pyrrolidine ring with the sugar on G22 and between the −OMe carbon C32 and the sugar on C11.

Water molecules play a significant role in stabilizing the drug−minor groove interactions. They fill the small gaps in the groove between each end of the TRIBIZ molecule and the deoxyribose sugars from neighboring molecules as part of the crystal packing. At the 5′ end and close to the methoxy group, a cluster of three water molecules has a largely space-filling role, hydrogen bonding to O2PC11 and O3′G10. At the pyrrolidine end of TRIBIZ the protonated nitrogen N1 makes a strong hydrogen bond to W37, which in turn interacts with W32 and W83 (Figure 7). These both make several strong hydrogen bonds, to N3 of G22 and N3 of A4 on the floor and to O4′ of C23 and O4′ of A5 on the walls of the groove.

DNA Conformation. All extended A_xT_x and $(AT)_x$ tract sequences exhibit high propeller twist angles for the AT base pairs, in both crystal and solution, both with native and with bound minor-groove drugs [for example, DiGabriele and Steitz (1993) and Coll et al. (1991)]. The present structure is no exception. Values for the propeller twist angles calculated by the CURVES program are given in Table 2, together with those for comparable crystal structures. Almost all of the AT base pairs in the present structure have propeller twist values >21°, in common with the native d(CGCAAATTTGCG)$_2$ structure (Edwards et al., 1992a) and the complex with distamycin (Coll et al., 1989). It is notable that the complexes with the smaller groove-binding molecules berenil and propamidine tend to have significantly lower propeller twists throughout the 5′-AAATTT tract.

There are several consequences of these high propeller twists. We note that the twists are about the adenine N1−thymine N3 axis, thus retaining the integrity of the N1−N3 Watson−Crick hydrogen bond but at the expense of the N6−O4 hydrogen bond (Table 3). The large propeller twist at A4·T21 brings the N6 atom of A4 into major-groove hydrogen bond contact with O4 of T20 and lengthens the A4N6−O4T21 distance, to somewhat greater than a standard hydrogen bond length. Similarly, the large twist at A5·T20 brings atom N6A5 into contact with O4T19 and lengthens the A5N6−O4T20 distance. The third and fourth AT base pairs follow this pattern, with the N6−O4 Watson−Crick hydrogen bond being weakened and three-center hydrogen bonding occurring in the major groove. The T8·A17 propeller twist does not appear to disrupt the T8O4−N6A17 hydrogen bond, although there is still a close major groove T8O4−N6A16 contact.

The large propeller twists result in values for interstrand distances between the hydrogen bond donor atoms of TRIBIZ and consecutive base pairs in the duplex that are significantly different from those for canonical B-DNA (Table 4a). Again, the short T7O2−T19O2 distance seen in the present structure is in accord with those found with the larger groove binders. The smaller molecules, such as berenil and propamidine, do not perturb this central region of the structure to the same extent.

The pattern of base-pair twist values in the present structure resembles those for the other d(CGCAAATTTGCG)$_2$ drug complexes, as well as the native structure (Table 4b), in all having a low value of ∼30° at the central A/T step. However, the total twist over the five A·T base pair steps is 14° less in the present structure than in the native one (Edwards et al., 1992a) and a consistent average of ∼10° less than in the others. This decrease in total helical twist as a result of TRIBIZ binding is equivalent to a local unwinding over the length of the tract.

DISCUSSION

The Nature of the Binding Site. The two drugs Hoechst 33258 and TRIBIZ both have a substituted benzene ring at one end of the molecule and a protonated, nonplanar saturated heterocyclic ring at the other. The linking benzimidazole moieties in both drugs bind to the 5′-AAATTT minor groove site in d(CGCAAATTTGCG)$_2$ by a combination of nonbonded van der Waals interactions to the (largely) hydrophobic walls (Spink et al., 1994; Vega et al., 1994) and bifurcated hydrogen bonds to N3 adenine and O2 thymine atoms on the floor of the groove. This behavior is paralleled in the several structures of Hoechst 33258 with the 5′-AATT-containing dodecamer (Pjura et al., 1987; Teng et al., 1988; Quintana et al., 1991) and the 5′-ATAT-containing one (Carrondo et al., 1989). In all of these

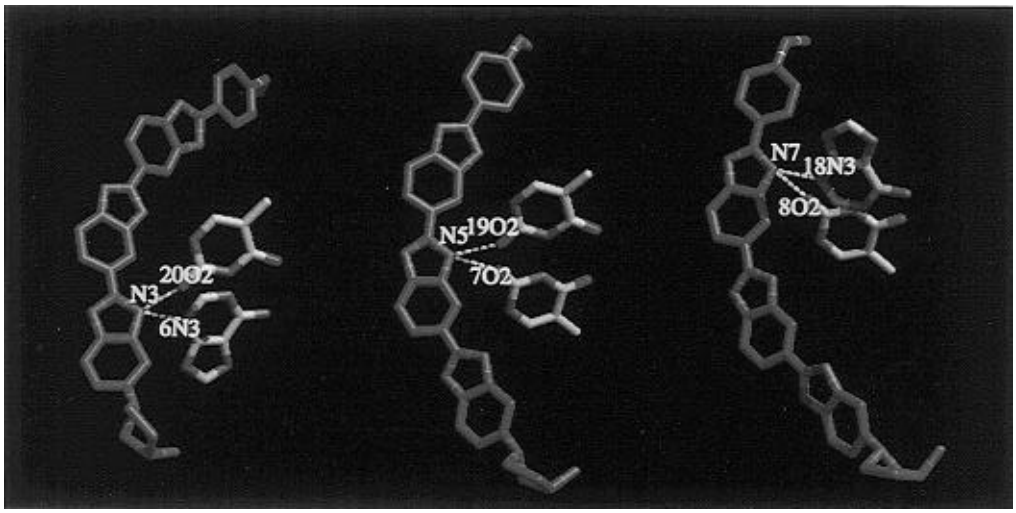

FIGURE 5: Plots of the hydrogen-bonding arrangements involving each benzimidazole group and the A/T base edges.

Table 1: Comparison of Hydrogen-Bonding Distances in Benzimidazole−d(CGCAAATTTGCG)$_2$ Complexes (in Å)

drug donor atom	DNA acceptor atom	Hoechst[a]	Hoechst[b]	TRIBIZ[c]
NH	O2(T8)	3.3	3.0	3.0
	N3(A18)	3.2	3.3	2.9
NH	O2(T7)	3.3	3.0	3.0
	O2(T19)	3.2	2.6	2.8
NH	N3(A6)			2.9
	O2(T20)			3.0

[a] Spink et al., 1994. [b] Vega et al., 1994. [c] This structure.

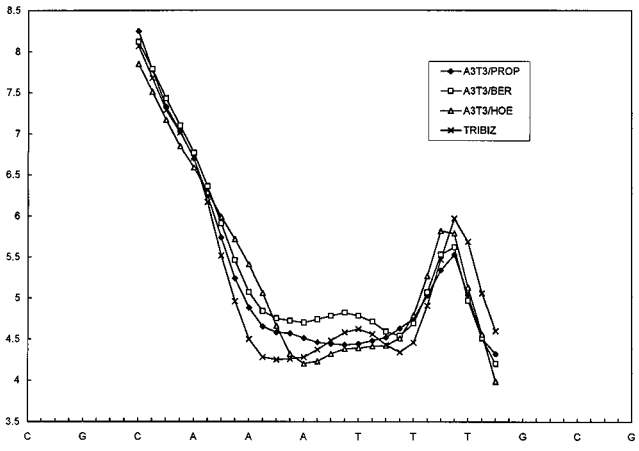

FIGURE 6: Minor groove width, based on C4′ separations, for the present structure and three other d(CGCAAATTTGCG)$_2$−drug complexes, calculated with the CURVES program (Lavery & Sklenar, 1992).

structures the bulky terminal N-heterocyclic ring (piperazine in Hoechst and pyrrolidine in TRIBIZ) is seen to require a widened groove for binding. This requirement governs the overall position of the drug along the sequence and thus defines which A/T bases along the AT tract can hydrogen bond to the benzimidazoles. Hoechst itself binds to the sequence 5′-ATTTG in its complex with d(CGCAAATTT-GCG)$_2$, with the wide minor groove accommodating the piperazine ring being at the 5′ end GC base pair. By contrast, the TRIBIZ molecule is oriented in the opposite direction (Figure 8), so that the equivalent pyrrolidine group is at the 3′ end of the binding site (which is best described by the sequence 5′-CAAATTTG). This orientation is optimal for formation of the three sets of benzimidazole bifurcated hydrogen bonds, so it appears that the TRIBIZ binding site

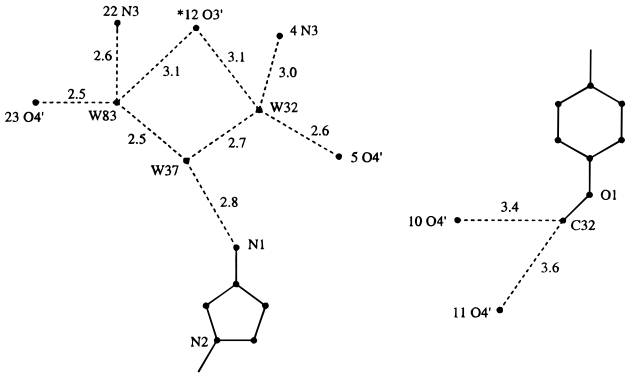

FIGURE 7: Schematic representation of the water arrangements and van der Waals contacts in the minor groove around the terminal pyrrolidine and methoxyphenyl groups of TRIBIZ.

Table 2: Propeller Twists (in deg), Calculated with the CURVES Program, for the Present Structure and Other Minor-Groove Drug Complexes with the Sequence d(CGCAAATTTGCG)$_2$, Using Coordinates Retrieved from the Nucleic Acid Database

	TRIBIZ[a]	A3T3[b]	HOE[c]	HOE[d]	DIST[e]	NET[f]	PROP[g]	BER[h]
C1−G24	−8	−23	−14	−21	−4	8	−9	−15
G2−C23	−11	−9	−1	−34	−2	−16	−6	−14
C3−G22	−9	−11	2	−18	−11	−1	−2	−6
A4−T21	−22	−21	−21	−23	−17	−3	−8	−17
A5−T20	−26	−22	−21	−23	−20	−18	−15	−25
A6−T19	−26	−24	−21	−23	−25	−27	−17	−19
T7−A18	−24	−21	−26	−29	−24	−14	−14	−19
T8−A17	−22	−25	−15	−31	−26	−25	−13	−21
T9−A16	−19	−22	−15	−28	−22	−17	−10	−15
G10−C15	−3	0	−7	−21	−8	−20	3	−9
C11−G14	−25	−22	−22	−28	−18	−31	−18	−25
G12−C13	5	10	−2	−16	−16	1	8	8

[a] Present work. [b] Native dodecamer (Edwards et al., 1992). [c] Hoechst 33258 (Spink et al., 1994). [d] Hoechst 33258 (Vega et al., 1994). [e] Distamycin (Coll et al., 1987). [f] Netropsin (Taberno et al., 1993). [g] Propamidine (Nunn & Neidle, 1995). [h] Berenil (Brown et al., 1992).

and orientation are governed both by hydrogen bonding and steric considerations. There are also interactions with the floor of the groove involving C2 atoms of adenine residues, as observed in other minor-groove drug−DNA complexes. However, in the present structure there are only two such contacts <3.5 Å.

Structural studies on the majority of complexes between oligonucleotides and ligands binding in the minor groove have shown that DNA conformation in the drug-free duplexes is little altered by bound drug, especially when the drug

Table 3: Interstrand Base-Pair Distances (in Å)

	classical Watson−Crick		bifurcated major groove
C1 N4−O6 G24	3.0		
C1 N3−N1	2.9		
C1 O2−N2	2.7		
G2 O6−N4 C23	2.6		
G2 N1−N3	2.7		
G2 N2−O2	2.6		
C3 N4−O6 G22	3.3		
C3 N3−N1	3.1		
C3 O2−N2	2.9		
A4 N6−O4 T19	3.2	A4 N6−O4 T20	3.4
A4 N1−N3	3.0		
A5 N6−O4 T18	3.4	A5 N6−O4 T19	3.4
A5 N1−N3	3.1		
A6 N6−O4 T17	3.4		
A6 N1−N3	2.9		
T7 O4−N6 A18	3.4	T7 O4−N6 A17	3.4
T7 N3−N1	2.9		
T8 O4−N6 A17	2.6	T8 O4−N6 A16	3.4
T8 N3−N1	2.6		
T9 O4−N6 A16	2.6		
T9 N3−N1	2.7		
G10 O6−N4 C15	2.9		
G10 N1−N3	2.7		
G10 N2−O2	2.5		
C11 N4−O6 G14	3.3		
C11 N3−N1	3.1		
C11 O2−N2	2.9		
G12 O6−N4 C13	2.5		
G12 N1−N3	2.7		
G12 N2−O2	2.8		

Table 4: Selected Geometric Parameters for the Present Structure and Other Drug Complexes with the Sequence $d(CGCAAATTTGCG)_2$[a]

(a) Interstrand Distances (in Å)[b]

	TRIBIZ	A3T3	HOE	HOE	DIST	NET	PROP	BER
6N3−20O2	3.8	4.0	3.9	3.7	3.4	3.4	3.8	3.6
7O2−19O2	3.4	4.0	3.5	3.4	3.7	3.8	4.2	3.8
8O2−18N3	3.8	3.9	3.8	3.6	3.6	4.6	4.0	4.0

(b) Base-Pair Twists (in deg) for the Five Steps of the A3T3 Tracts in the Various Structures[c]

	TRIBIZ	A3T3	HOE	HOE	DIST	NET	PROP	BER
step 1	35	35	38	32	40	34	34	37
step 2	35	40	40	42	34	40	40	38
step 3	30	32	30	30	36	29	31	32
step 4	35	39	37	37	32	38	33	37
step 5	35	38	38	37	39	41	42	38
total	170	184	183	178	181	182	180	182

[a] Structures are as in Table 2. [b] Values of corresponding N3−O2 and O2−O2 interstrand distances in canonical B-DNA are 4.6 and 4.8 Å, respectively. [c] The value for canonical B-DNA is 36°.

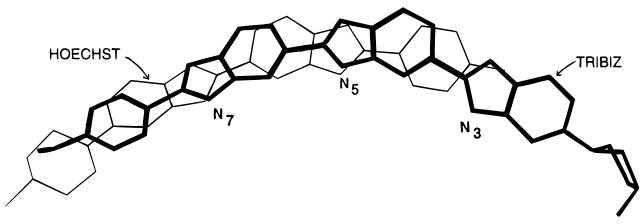

FIGURE 8: Overlay of the structures of the TRIBIZ and Hoechst 33258 (in bold) molecules as seen in the present crystal structure and that determined by Spink et al. (1994). Atom labels are those used for TRIBIZ.

molecule, and hence its binding site, is small. Hydrogen bonding appears to be secondary to stabilization from van der Waals interactions, with the binding site predetermined by optimal fitting to a particular region of appropriate groove width. With TRIBIZ, we see a distinct picture. First, groove width is significantly decreased for much of the binding site. Second, there are a number of small but necessary changes to DNA geometry in order to ensure that the drug is in optimal register with the hydrogen-bonding groups on the DNA. Thus, although the rms difference between the 5′-AAATTT site in native (Edwards et al., 1992a) and TRIBIZ-bound crystal structures is only 0.34 Å, this small overall difference masks the many local differences in propeller and helical twists outlined above. The changes in helical twist suggest that TRIBIZ binding is also accompanied by a degree of local helix unwinding.

Phasing and Isohelicity of Drug Binding. The ability of a DNA-binding molecule to make geometrically optimal hydrogen bonds with donor/acceptor groups on successive base-pair edges of a DNA can be a key factor in overall effective recognition. In the case of small ligands such as berenil and propamidine, two other factors are likely to be of greater importance: (1) the shape complementarity between the concave drug surface and the convex surface of the minor groove floor and (2) effective drug−DNA nonbonded contacts with the walls of the groove. Factor 1 can be defined by measures of isohelicity (Goodsell & Dickerson, 1986). This shape complementarity can be approximated by the matching of two-dimensional circular radii (with the center at the helix axis and with one radius of curvature for DNA and one for drug) when considering such small ligands or individual subunits of larger ones. However, once several subunits become strung together, then shape complementarity becomes more complex since such molecules require an inherent twist in order to achieve and maintain hydrogen bonding for each subunit. We see here such a twist for TRIBIZ, of ∼50°, and as a consequence it is no longer possible to describe isohelicity in simple "matching radii" terms. When this twist is reduced to 0°, it is not surprising that some of the hydrogen bonds between benzimidazole units and the A·T base pairs do not appear to be of acceptable geometry.

Structural studies on minor-groove drug complexes have shown that isohelicity is optimal when there are a series of close nonbonded contacts between C2 atoms of adenines and atoms representing the full extent of the inner concave surface of a drug molecule (Kopka et al., 1985; Brown et al., 1992). We see that this is not the case for the present TRIBIZ complex, with only the methoxyphenyl ring and benzimidazole subunit 3 having such contacts. The TRIBIZ molecule thus does not have an effective isohelical fit to this particular sequence. If it were to do so, with close contacts to the groove floor involving subunits 1 and 2 as well, then the pattern of hydrogen bonds between base-pair edges and imidazole nitrogens (Figures 4 and 5) could no longer be maintained. This in turn suggests that isohelicity is of lesser importance for TRIBIZ (and implicitly for other large minor-groove binders) than effective phasing of subunits with successive base-pair edges, together with the maintainance of close van der Waals contacts between the whole length of the drug and atoms on the walls of the minor groove.

The other factor in effective phasing that this study highlights is the structure of the binding site itself. The 5′-AAATTT central sequence in this structure has a rms deviation of 2.5 Å when compared to that in canonical DNA. The N3−O2 and O2−O2 cross-strand distances critical for TRIBIZ hydrogen bonding are up to 1.4 Å longer in a

canonical DNA structure (Table 4a), and an acceptable molecular model for TRIBIZ with canonical DNA that has equivalent hydrogen bonding to that found here cannot be constructed. This is in accord with earlier suggestions from modeling studies with canonical B-DNA that the benzimidazole subunit is ~20% too long for effective isohelicity (Goodsell & Dickerson, 1986), even though a more mathematical treatment taking helical twist parameters into account (Zasedatelev, 1991) suggested that a benzimidazole trimer (as is present in TRIBIZ) approaches a reasonable fit to the minor groove of canonical B-DNA. The present structure suggests that the TRIBIZ molecule achieves optimal phased hydrogen bonding between successive sites along the sequence by only binding to particular sequences, exemplified by 5′-AAATTT, that are in principle sufficiently deformable from the canonical structure. The observations in several crystal structures that this tract is always significantly altered in structure from that of a canonical sequence further suggest that predictions of appropriate subunits and linkers for building up molecules capable of effectively recognizing DNA sequences greater than 7−8 bases pairs in length need to take into account the detailed structure and flexibility of the sequence involved, rather than solely considerations of optimal steric fit.

REFERENCES

Abu-Daya, A., Brown, P. M., & Fox, K. R. (1995) *Nucleic Acids Res. 23*, 3385−3392.

Bathini, Y., Ekambareswara, R., Shea, R. G., & Lown, J. W. (1990) *Chem. Res. Toxicol. 3*, 268−280.

Beerman, T. A., McHugh, M. M., Sigmund, R., Lown, J. W., Rao, K. E., & Bathini, Y. (1992) *Biochim. Biophys. Acta 1131*, 53−61.

Bellorini, M., Moncollin, V., D'Incalci, M., Mongelli, N., & Mantovani, R. (1995) *Nucleic Acids Res. 10*, 1657−1663.

Boykin, D. W., Kumar, A., Spychala, J., Zhou, M., Lombardy, R. J., Wilson, W. D., Dykstra, C. C., Jones, S. K., Hall, J. W., Tidwell, R. R., Laughton, C., Nunn, C. M., & Neidle, S. (1995) *J. Med. Chem. 38*, 912−916.

Broggini, M., Erba, E., Ponti, M., Ballinari, D., Geroni, C., Spreafico, F., & D'Incalci, M. (1991) *Cancer Res. 51*, 199−204.

Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I., & Neidle, S. (1990) *EMBO J. 9*, 1329−1334.

Brown, D. G., Sanderson, M. R., Garman, E., & Neidle, S. (1992) *J. Mol. Biol. 226*, 481−490.

Brünger, A. T., Kuriyan, J., & Karplus, J. (1987) *Science 235*, 458−460.

Carrondo, M. A. A. F. de C. T., Coll, M., Aymami, J., Wang, A. H.-J., van der Marel, G. A.,, van Boom, J. H., & Rich, A. (1989) *Biochemistry 28*, 7849−7859.

Chen, A. Y., Chiang, Y., Gatto, B., & Liu, L. F. (1993) *Proc. Natl. Acad. Sci.* U.S.A. *90*, 8131−8135.

Chen, Y.-H., & Lown, J. W. (1994) *J. Am. Chem. Soc. 116*, 6996−7005.

Chiang, S,-Y., Welch, J., Rauscher, F. J., III, & Beerman T. A. (1994) *Biochemistry 33*, 7033−7040.

Coll, J., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385−8389.

Cory, M. J., Tidwell, R. J., & Fairley, T. A. (1992) *J. Med. Chem. 35*, 431−438.

Dervan, P. B. (1986) *Science 232*, 464−471.

DiGabriele, A. D., & Steitz, T. A. (1993) *J. Mol. Biol. 231*, 1024−1039.

Dorn, A., Affolter, M., Müller, M., Gehring, W. J., & Leupin, W. (1992) *EMBO. J. 11*, 279−286.

Edwards, K. J., Brown, D. G., Spink, N., Skelly, J. V., & Neidle, S. (1992a) *J. Mol. Biol. 226*, 1161−1172.

Edwards, K. J., Jenkins, T. C., & Neidle, S. (1992b) *Biochemistry 31*, 7104−7109.

Fede, A., Labhardt, A., Bannwarth, W., & Leupin, W. (1991) *Biochemistry 30*, 11377−11388.

Fede, A., Billeter, M., Leupin, W., & Wüthrich, K. (1993) *Structure 1*, 177−186.

Feng, J. A., Johnson, R. C., & Dickerson, R. E. (1994) *Science 163*, 348−355.

Ferrin, T. E., Huang, C. C., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.

Goodsell, D., & Dickerson, R. E. (1986) *J. Med. Chem 29*, 727−733.

Gravatt, G. L., Baguley, B. C., Wilson, W. R., & Denny, W. A. (1994) *J. Med. Chem 37*, 4338−4345.

Harshman, K. D., & Dervan, P. B. (1985) *Nucleic Acids Res. 13*, 4825−4835.

Henderson, D., & Hurley, L. H. (1995) *Nat. Med. 1*, 525−527.

Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E., & Langridge, R. (1991) *J. Mol. Graphics 9*, 230−236.

Kopka, M. J., & Larsen, T. A. (1992) *Nucleic Acid Targeted Drug Design* (Propst, C. L., & Perun, T. J., Eds.) pp 304−374, Marcel Dekker, Inc., New York.

Kopka, M. J., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *J. Mol. Biol. 183*, 553−563.

Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.

Mrksich, M., Parks, M. E., & Dervan, P. B. (1994) *J. Am. Chem. Soc. 116*, 7983−7988.

Nunn, C. M., & Neidle, S. (1995) *J. Med. Chem. 38*, 2317−2325.

Patel, S. R., Kvols, L. K., Rubin, J., O'Connell, M. J., Edmonson, J. H., Ames, M. M., & Kovach, J. S. (1991) *Invest. New Drugs 9*, 53−57.

Perun, T. J. (1992) *Nucleic Acid Targeted Drug Design* (Propst, C. L., Ed.) pp 303−373, Marcel Dekker, Inc., New York.

Pjura, P. E., Grzeskowiak, K., & Dickerson, R. E. (1987) *J. Mol. Biol. 197*, 257−271.

Prakash, A. S., Valu, K. K., Wakelin, L. P. G., Woodgate, P. D., & Denny, W. A. (1991) *Anti-Cancer Drug Des. 6*, 195−206.

Quintana, J. R., Lipanov, A. A., & Dickerson, R. E. (1991) *Biochemistry 30*, 10294−10306.

Singh, M. P., Plouvier, B., Hill, C. G., Gueck, J., Pon, R. T., & Lown, J. W. (1994) *J. Am. Chem. Soc. 116*, 7006−7020.

Spink, N., Brown, D. G., Skelly, J. V., & Neidle, S. (1994) *Nucleic Acids Res. 22*, 1607−1612.

Sun, Q., Gatto, B., Yu, C., Liu, A., Liu, L. F., & LaVoie, E. J. (1994) *Bioorg. Med. Chem. Lett. 4*, 2871−2876.

Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G. A., van Boom, J. H., Rich, A., & Aymami, J. (1993) *Biochemistry 32*, 13838−13843.

Teng, M., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671−2690.

Tidwell, R. R., Jones, S. K., Geratz, J. D., Ohemeng, K. A., Cory, M., & Hall, J. E. (1990) *J. Med. Chem. 33*, 1252−1257.

Vega, M. C., Saez, I. G., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., & Coll, M. (1994) *Eur. J. Biochem. 222*, 721−726.

Welch, J. J., Rauscher, F. J., III, & Beerman, T. A. (1994) *J. Biol. Chem. 269*, 31051−31058.

Werner, M., Huth, J., Gronenborn, A., & Clore, G. M. (1995) *Cell 81*, 705−714.

Wong, S. S. C., Sturm, R. A., Michel, J., Zhang, X,-M., Danoy, P. A. C., McGregor, K., Jacobs, J. J., Kaushal, A., Dong, Y., Dunn, I. S., & Parsons, P. G. (1994) *Biochem. Pharmacol. 46*, 827−837.

Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D. W., & Neidle, S. (1995) *Nucleic Acids Res. 23*, 3678−3684.

Youngquist, R. S., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2565−2569.

Youngquist, R. S., & Dervan, P. B. (1987) *J. Am. Chem. Soc 109*, 7564−7566.

Zasedatelev, A. S. (1991) *FEBS Lett. 28*, 209−211.

Zimmer, C., & Wähnert, U. (1966) *Prog. Biophys. Mol. Biol. 47*, 31−112.

BI960421M